

using a computer to perform a fuzzy similarity join on the database to identify at least one item from the database that comprises a property similar to a property of the target item.

Sub
El
AI
Conf.

11. A method according to claim 10 wherein the fuzzy similarity join is a chemical similarity join, and the target item and identified item are chemical compounds.

12. A method according to claim 10 wherein a user of the computer is informed of the identification of one or more identified items that comprise a property similar to the property of the target item.

13. A method according to claim 11 wherein the property comprises a chemical structure of the target item.

14. A method according to claim 10 wherein a plurality of properties of the target item are identified.

15. A method according to claim 11 wherein a plurality of properties of the target item are identified.

16. A method according to claim 15 wherein one or more properties of the plurality of properties is selected from the group consisting of chemical structure, synthesis pathway, binding data, biological activity, structure-activity relationship information, molecular weight, partition coefficient, electric charge, size, efficacy, toxicology, manufacturer, price, and availability.

17. A method according to claim 12 wherein the user is informed via a remote communication link.

18. A method according to claim 17 wherein the remote communication link is the Internet.

SUB E3
419. A method according to claim 11 further comprising eliminating test items from the database by selection of user-defined criteria for non-desired items.

420. A method according to claim 10 wherein the target item is a biological compound.

421. A method according to claim 20 wherein the biological compound is a protein.

SUB E3
422. A method according to claim 21 wherein the target item is a gene.

SUB E3
423. A computer-based method for identifying, from a database comprising chemical compounds, at least one chemical compound having at least one property similar to a target chemical compound, the method comprising:

identifying a property of a target chemical compound; and

using a computer to perform a chemical similarity join on the database to identify at least one database chemical compound that has a property similar to the property of the target chemical compound.

SUB E3
424. A method according to claim 23 wherein a user is informed of the identification of at least one database chemical compound that has a property similar to the property of the target chemical compound.

SUB E3
425. A method according to claim 23 wherein the property of the target chemical compound comprises the chemical structure of the target chemical compound.

SUB E3
426. A method according to claim 23 wherein a plurality of properties of the target chemical compound are identified.

427. A method according to claim 23 wherein the identified property is a neighborhood effect.

428. A method according to claim 27 wherein the neighborhood effect comprises a range of values of property metrics for the target chemical compound.

SUB E6
429. A method according to claim 23 wherein the similarity between the properties of the target chemical compound and the database chemical compound is determined using at least one parameter selected from the group consisting of a Tanimoto coefficient and a Molecular hologram.

SUB H/FI
430. A method according to claim 24 wherein the user is informed via a remote communication link.

431. A method according to claim 30 wherein the remote communication link is the Internet.

SUB E7
432. A method according to claim 23 further comprising excluding at least one database chemical compound from chemical similarity join by selecting user-defined exclusion criteria for non-desirable compound features.

433. A method according to claim 23 wherein the chemical property is selected from the group consisting of chemical structure, synthesis pathway, binding data, biological activity, structure-activity relationship information, molecular weight, partition coefficient, electric charge, size, efficacy, toxicology, manufacturer, price, and availability.

434. A computer-based method for identifying, from a database comprising biological compounds, at least one biological compound having at least one property similar to a target biological compound, the method comprising:

identifying a property of a target biological compound; and
using a computer to perform a fuzzy similarity join on the database to identify at least one database biological compound that has a property similar to the property of the target biological compound.

435. A method according to claim 34 wherein the biological compound is a protein.

436. A method according to claim 34 wherein the biological compound is a gene.